



*Citation for published version:*

Hudson, J 2012 'Internet Banking and the Marginal Internet User' Bath Economics Research Working Papers, no. 7/12, Department of Economics, University of Bath, Bath, U. K.

*Publication date:*  
2012

*Document Version*  
Publisher's PDF, also known as Version of record

[Link to publication](#)

© The Author

**University of Bath**

**Alternative formats**

If you require this document in an alternative format, please contact:  
[openaccess@bath.ac.uk](mailto:openaccess@bath.ac.uk)

**General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

**Take down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Internet Banking and the Marginal Internet User

John Hudson

No. 7/12

**BATH ECONOMICS RESEARCH PAPERS**

**Department of Economics**

Department of  
Economics



UNIVERSITY OF  
**BATH**

# Internet Banking and the Marginal Internet User

John Hudson\*

## *ABSTRACT*

\*Department of Economics, University of Bath, Bath, BA2 7AY, UK, email:  
[J.R.Hudson@bath.ac.uk](mailto:J.R.Hudson@bath.ac.uk)

### Abstract

Using Heckman's sample selection methodology to analyse individual use of Internet banking we find a negatively significant Inverse Mills ratio. This is consistent with other studies. We go on to argue that this indicates that the probability of using Internet banking declines with the marginality of the Internet user. We further link this marginality to low cost Internet access and thence trust in this access. The inverse Mills ratio squared is also significant, thus both questioning the assumption of a bivariate normal distribution between the error terms from the selection and Internet banking equations, and also suggesting that there is a nonlinear relationship between Internet user marginality and the probability of using it for Internet banking.

Key Words: Marginal user, sample selection, Internet banking.

JEL Classification: C24, G21.

# Internet Banking and the Marginal Internet User

## 1 Introduction

In recent years there has been a substantial growth in Internet banking across many countries. There are, however, wide variations between countries in the proportion of Internet users who use online banking. According to the EU, usage in 2011 was highest in the Scandinavian and Baltic countries at around 90% for Norway, Finland and Estonia and lowest for the southern periphery of Europe at less than 20% for Bulgaria, Romania and Greece<sup>1</sup>. Internet banking has the potential to increase efficiency in firms and wellbeing in households, particular within peripheral locations remote from traditional banking facilities. It is also a crucial aspect of e-commerce. Thus it is of importance to understand the factors which lead some countries to lag behind others in the adoption of Internet banking and why in all countries there are a substantial proportion of the population who do not use it.

It is an example of the diffusion of a new technology and, as with other examples of new technology, diffusion meets resistance from people unaware of how to use the technology and who to an extent, mistrust that technology. In the case of Internet banking such mistrust may be justified. According to the European Commission (2010) there is a growing incidence of identity theft and online fraud, with attacks becoming increasingly sophisticated (trojans, botnets, phishing, malware, spyware, and password sniffing). Other factors which may influence individual usage of Internet banking include 'need', which tends to vary by socio-economic characteristics. The ones which have found to be important in the literature include age, education, gender, location, income and marital status.

In this paper our focus is not so much on the socio-economic determinants of Internet banking, nor on the substantial differences between countries, but on the impact on accessing

---

<sup>1</sup> [http://ec.europa.eu/information\\_society/digital-agenda/scoreboard/graphs/index\\_en.htm](http://ec.europa.eu/information_society/digital-agenda/scoreboard/graphs/index_en.htm)

Internet banking of being a marginal Internet user, which we link to perceived risk of, and trust in, Internet banking and the Internet itself. The specific question we wish to pursue is, given their socio-economic variables as they appear in the Internet banking equation, are all people who have access to the Internet equally likely to use it for Internet banking? The hypothesis we put forward is that the probability of using Internet banking declines with the marginality of being an Internet user. A marginal Internet user is defined as one with a low probability of using the Internet based on their observed socio-economic characteristics. We will identify the marginality of being an Internet user through the use of Heckman's (1976, 1979) sample selection methodology. The work done analysing individual use of Internet banking has tended to rely either on an analysis of survey responses by users, or potential users (Gerrard et al, 2006, Cunningham et al, 2005), or econometric analysis. The latter is generally based on Heckman's methodology (see for example Lee et al, 2004 and Orviska and Hudson, 2009). We pursue the latter approach and as with other papers use Heckman's methodology to deal with sample selection bias, but also to help in measuring Internet user marginality.

The paper proceeds as follows. In the next section we review the literature. We then set out the methodology discussing the use of Heckman's sample selection model and the role of the Inverse Mills ratio. We will argue that the Inverse Mills ratio can be viewed as reflecting the marginality of the Internet user which we in turn link to confidence in the Internet. This is an assumption we go some way to justifying. It is also consistent with the literature, where those who use Heckman's sample selection methodology to analyse Internet banking tend to find a significantly negative inverse Mills ratio (Lee et al, 2004, Orviska and Hudson, 2009). This is also the case with the related area of e-commerce (Goldfarb and Prince, 2008). In the next section we discuss the data which is based on several Eurobarometer studies. We then present the results and finally conclude the paper. In presenting the results, we argue that the full impact of a variable which appears in both equations of the Heckman procedure needs to take account of an indirect impact via the inverse Mills ratio.

## **2 Literature Review**

### **2.1 Internet banking**

The purpose of this section is to identify the variables, in particular the control variables, to include in the analysis. The spread of Internet banking is, as we have already emphasised, an example of the diffusion of a new technology, and to an extent, diffusion occurs by people learning from others. Hence diffusion can be expected to be most rapid in central, urban areas (Furst et al 2002). The quality of Internet connection is also often better in urban areas. On the other hand, people in less well populated areas, remote from bank branches may have more to gain from Internet banking, given they already have access to the Internet. There is some evidence on this, although it is not totally consistent. Thus, in the USA, even where broadband is available in rural areas, adoption has lagged behind that in urban areas (La Rose et al, 2007). However, Sinai and Waldfogel (2004) in a study of the USA also find evidence that individuals connect to the Internet to overcome local isolation. Rural adoption would be one example of 'need'. Another relates to the extent of individual savings and financial transactions, which are likely to be greater for richer people and also for older and more educated people, as both the latter tend to be positively linked with wealth. Partly because of such linkages, most, although not all, studies identify socio-economic factors such as education and gender to be significant determinants of Internet banking (Guerrero et al., 2007). Courchane et al. (2002) and De Young et al. (2007) also indicate that take up will be greater amongst the young and the better educated. Orviska and Hudson (2009) identify an inverted U shaped relationship linking the probability of using Internet banking to age, with the turning point coming in the mid 50s. Regions characterized by high per capita income (Sullivan and Wang, 2005), employment growth (DeYoung et al., 2007) and Internet literacy (Bughin, 2003) have also been listed as factors determining the diffusion of Internet banking.

In terms of the individual rationale for the non-adoption of Internet banking, security fears are important (Cox and Dale, 2001; Howcroft et al., 2002; Lee et al, 2004 and Orviska and Hudson, 2009). Cheng et al. (2006) cite, with reference to Hong Kong, ease of use and fears

over security. Security and convenience are also prominent in explaining the adoption of Internet banking in Malaysia (Poon, 2008) and the EU (Guerrero et al., 2007). Security risk is closely related to ‘trust,’ which has been the specific focus of a paper by Kelton et al. (2008). Trust and perceived risk are seen as major factors in determining peoples use of Internet banking specifically and e-commerce more generally (Dimitriadis and Kyrezis, 2011; Aldas-Manzano et al, 2011, Mukherjee and Nath, 2003; Wang et al., 2003). However the focus tends to be on trust in the Web service provided by the bank, rather than the Internet or the Internet provider per se. But there have been some studies on other aspects of Internet usage, particularly for e-Commerce, which focus on the different kinds of risk and trust, and in particular emphasise trust in the Internet per se (Lee and Turban, 2001). McKnight et al. (2002) emphasise both trust in the Web provider, e.g. the online bank, and the Internet environment itself. They emphasise four elements of trust: trust in the safeguards of the Internet, trust in the legal and technological structures of the Internet, trust in devices such as encryption to provide protection various threats and trust in the robustness and safety of the Internet environment. The studies tend to be of two kinds, first evaluation of surveys of users and secondly econometric analysis of Internet usage. Trust and risk aversion may be proxied by information from survey questions, but will also vary by socio-economic characteristics. For example on several dimensions women have been shown to be more risk averse than men (Jianakoplos and Bernasek, 1998)

## 2.2 Internet usage

Orviska and Hudson (2009) linked internet usage to age, population density, through a rural urban distinction, employment status, occupation, education and marital status. Goldfarb and Prince (2008) linked Internet adoption to a similar set of variables, including age marital status, city dwellers, education and income, but concluded that it did not depend upon gender nor the number of children. More generally, the literature suggests that within a given country, socio-economic status is critical. The “digital divide” has been defined as the phenomenon where people of higher socio-economic status demonstrate greater access and

usage compared to those from lower socio-economic status groups (DiMaggio et al, 2001). Recabarren et al. (2007) show that in part Internet knowledge and skills are dependent upon the type of school a person attended, and that individuals belonging to a subculture of low social strata do not have the knowledge necessary to make effective use of the Internet. This reflects both the technical ability to be able to use the Internet and the potential advantages of doing so in terms of the greater range of services that a wealthier person can benefit from. But in addition, the cost of accessing the Internet is linked to the cost, or even the physical possibility, of accessing broadband (Kontos et al. 2007). Access is either via a dial up link or over broadband. Both make use of a fixed phone link, but the former is much slower, and hence usage of the Internet more problematic and of less net benefit to the user. Broadband access is linked to location, and in general people in rural areas are least well served, with access to only low quality broadband, and even in some cases no access at all. La Rose et al. (2007) argue that in the USA there were initially gender differences in Internet usage but that these have disappeared over time. They also associate income and age with Internet usage.

Although many of the studies are country specific, there has been some analysis of differences between countries. Chinn and Fairlie (2006), using a panel of 161 countries over the 1999-2001 period, conclude that significant explanatory factors included electricity power consumption, the youth dependency ratio, urban population, per capita income and regulatory quality. The significance of the latter indicates the importance of the institutional environment. Human capital is not significant. The urban population ratio is negatively significant. That result again suggests that the Internet compensates for peripherality and thus can help reduce disparities between rural and urban regions. Kiiski and Pohjola (2002) also consider Internet diffusion across countries and find GDP per capita and Internet cost are significant in explaining Internet growth in a sample of OECD countries. In a wider sample of countries, investment in education also becomes significant. Hargittai (1999) in an early study of telecommunications policy also concludes that amongst OECD countries, economic wealth is one of the most significant factors in determining Internet connectivity. Also important is the telecommunications policy of the country.



### 3 The sample selection model

Someone will use the Internet if the gains outweigh the costs. A marginal user will be one for whom either the benefits are greater than one would expect given their socio-economic characteristics, or the costs are less than would have been expected. Positive benefits may arise for reasons unique to the individual, such as a desire for knowledge or distant friends who can most economically be communicated to electronically. Such unexpected benefits should not impact upon the probability of using Internet banking and thus we focus on lower costs. The costs of Internet provision vary between providers and also for the same provider. For example one leading British web site provider offers increasingly expensive packages offering both faster speeds and enhanced security. The literature suggests that both will impact on the likelihood of using Internet banking. Differential costs may also arise because of shared access with others or the person has access to the Internet via Internet cafes, or some freely available wireless connection as in some hotels. In all these cases security may be a cause for concern when compared to access via a single user home connection, and speed too may be a factor. Speed of Internet use may also be a factor impact for rural Internet users. From the literature review both these factors are known to impact on individual use of Internet banking. Hence we would expect marginal Internet users to be less likely to use Internet banking.

How are we to include the marginal Internet user in our analysis? In this research we make use of Heckman's sample selection methodology. This is a two stage technique. In the first stage Internet usage is analysed using a binomial probit model. The Inverse Mills ratio is then calculated and used as an additional explanatory variable in a second stage estimation of Internet banking usage, again using a probit model. The Inverse Mills ratio is the ratio of the probability density function  $\phi(-\mathbf{Z}_i'\boldsymbol{\gamma})$  to the cumulative density function  $(\Phi(\mathbf{Z}_i'\boldsymbol{\gamma}))$ , where  $\mathbf{Z}_i'\boldsymbol{\gamma}$  is the predicted value from the first stage regression based on a vector of explanatory variables,  $\mathbf{Z}_i$ . The Inverse Mills ratio is this a monotone decreasing function of the probability that an observation is selected into the sample, and therefore an increasing function of the probability

that it is not selected into the sample ( $\Phi(-\mathbf{Z}_i'\boldsymbol{\gamma})$ ).  $\varepsilon_{ij}$  and  $v_i$ , the error terms from the two equations. In the standard Heckman methodology the assumption is made that they follow a bivariate normal distribution. If this is not correct then simply entering the inverse Mills ratio in the second stage equation will not fully correct for sample selection bias.

The inclusion of the Inverse Mills ratio is generally regarded as correcting for the statistical problems raised by sample selection bias. But, it can be also regarded as the inclusion of an additional variable. This is what Kai and Nagpurnanand (2007) argue. They were analysing, within the context of corporate finance, the choice of investment bank to underwrite a security issue. This is, of course, conditional upon issuing such an issue in the first place. This was modelled as a function of publicly available information and the error term was regarded as reflecting private information driving the corporate decision to be modelled. In the context of the wage equation, e.g., it could reflect information on the individual's reservation wage. In the context of our analysis, the Inverse Mills ratio could also be interpreted as private factors influencing the decision to adopt the Internet. Alternatively, and this is the interpretation we adopt, it can be regarded as reflecting the marginality of the decision. The inverse Mills ratio, as already emphasised, is inversely related to the probability  $\Phi(\mathbf{Z}_i'\boldsymbol{\gamma})$  that an observation is selected into the sample. A high value for the inverse Mills ratio means the probability of being in the sample is low, i.e. in the context of Internet banking they are marginal Internet users<sup>2</sup>.

From this perspective, bias then arises because some of the right hand side variables in the Internet banking equation are correlated with the marginality of being an Internet user. This interpretation of the Inverse Mills ratio then immediately raises an additional issue, although one related to the assumption that the error terms from the two error terms follow a bivariate normal distribution. It is a semi-continuous variable with a lower bound of 0. Simply focusing on this as a variable in its own right, why, as a semi-continuous variable, should its impact on the decision to adopt Internet banking be linear? Why should an increasing degree of

---

<sup>2</sup> Alternatively in the context of the labor market, they would be marginal entrants, which may suggest they have a relatively low reservation wage.

marginality of the decision to adopt the Internet have a linear impact on the decision to use Internet banking? This is an issue which has not really been explored in the literature, although Vella (1998) did compare the standard approach with including the predicted value and its square from the selection equation in the second stage equation, concluding the two approaches were virtually equivalent. In addition, Pagan and Vella (1989) add on higher order terms  $(\mathbf{Z}_i'\boldsymbol{\gamma})^j[(\phi(\mathbf{Z}_i'\boldsymbol{\gamma})/(1-\Phi(-\mathbf{Z}_i'\boldsymbol{\gamma})))]$  proportional to the Inverse Mills ratio, shown in [.]. The joint significance of these additional terms is then a test for the joint normality of the error terms. Johnston and Dinardo (1997) also suggest including as additional regressors the squared value of the inverse Mills ratio.

The specification of the two equations, the first stage, Internet user one and the second stage Internet banking one, are as suggested by the literature, in particular Orviska and Hudson (2009). The explanatory variables will include both socio-economic variables and variables related to trust. In order to estimate the model it is desirable that there are some variables included in the first stage of the analysis which are excluded from the second. In our analysis education, marital status and employment status fulfill this function, as well as certain other variables. The assumption, which was confirmed by the data, with respect to the first two in particular is that if an individual has the knowledge to become connected to and then use the Internet, then they also have the knowledge to use the Internet for the purpose of on-line banking. Being unemployed puts financial pressure on the individual, but having incurred the expense of connecting to the Internet, there is no further expense involved with using it for on-line banking.

The two step estimator, i.e. estimating the two equations separately with the inverse Mills ratio calculated from the first stage regression entered into the second stage regression as an explanatory variable, is not efficient because the variance of the error term from the second stage regression is not homoscedastic. Heckman dealt with this by calculating the asymptotic covariance of the two step estimator, given the assumptions he made concerning the relationship between  $\varepsilon_{ij}$  and  $v_{i1}$ , in particular that their joint density function is bivariate

normal. It is also possible to use a covariance estimator such as White's (1980) to deal with heteroscedasticity (Frances and Paap, 2004), and this is the approach we adopt.

#### **4 The data and empirical specification**

We will be using data from the Eurobarometer surveys carried out in October/December 2004 (Eurobarometer 62.1) and October/November 2005 (Eurobarometer 64.2) of the EU member countries and a more recent data from a survey carried out in November/December 2010 (Eurobarometer 74.3). The earlier two surveys were analysed in Orviska and Hudson (2009), but there the focus was on the socio-economic determinants of several types of Internet usage. Here the focus is specifically on Internet banking and the impact on this of marginal Internet users. The surveys cover the populations of the respective nationalities of the European Union member states aged fifteen years and over, although we restrict our analysis to those aged between 18 and 97. Later surveys also included data on Bulgaria, Romania, Croatia, Turkey and Northern Cyprus, although to facilitate comparability these countries were not included in the regression sample. The Eurobarometer surveys are carried out by TNS Opinion and Social Consortium at the request of the European Commission. The basic sample design is a multi-stage, random probability one. The surveys are designed to be representative in terms of the distribution of the resident population of the respective EU nationalities with respect to metropolitan, urban and rural areas. All interviews are face to face, in people's homes and in the appropriate national language. All the variables, including the socio-economic ones, are defined in an appendix at the end of this paper.

The dependent variables in our regression analysis are measures of Internet usage for banking and Internet access. Independent variables from the Eurobarometer data set were as defined in Orviska and Hudson (2009) and include age, a measure of education, gender, location and employment status. The 2009 data set contains information on trust in banks, which we use in the second stage regression, and individual financial hardship. The 2005 data set contains information on the individual respondent's trust in business which is also used in the second stage regression. The 2004 data set has information on the quality of the

respondent's fixed telephone line, which proxies the quality of their Internet link. We also use country dummy variables.

## 5 Regression analysis

*Insert Table 1 about here.*

The results are shown in Table 1. The first five columns relate to the earliest survey. Column 1 shows the results using the heckprob command in STATA. Column 2 replicates these results using the two stage estimator based on successive probit regressions on the selection and usage equations, Standard errors have been corrected for using the robust or sandwich estimator to deal with problems of heteroscedasticity. The two sets of results are very similar. There are some minor differences as heckprob uses a maximum likelihood methodology to estimate both equations jointly. If we add  $(\mathbf{Z}_i'\boldsymbol{\gamma})(\phi(\mathbf{Z}_i'\boldsymbol{\gamma})/(1-\Phi(-\mathbf{Z}_i'\boldsymbol{\gamma})))$  and  $(\mathbf{Z}_i'\boldsymbol{\gamma})^2(\phi(\mathbf{Z}_i'\boldsymbol{\gamma})/(1-\Phi(-\mathbf{Z}_i'\boldsymbol{\gamma})))$  to the regression then these are jointly significant at the 1% level, thus, as suggested by Pagan and Vella (1989), rejecting the hypothesis of joint normality between the error terms. This was also the case in the other two sets of regressions we discuss later and suggests the inverse Mills ratio is not in itself sufficient to solve the problem of sample selection bias. However, these additional regressors do not in themselves provide any ready intuition with respect to the impact of Internet user marginality, which is our prime concern. Hence, the third column adds to the second stage regression the square of the inverse Mills ratio. It is highly significant at the 1% level of significance. Figure 1 shows the kernel density<sup>3</sup> of the Inverse Mills ratio for those with access to the Internet, i.e. the sample in the second stage regression. It also shows the combined impact on the probability of using Internet banking from the inverse Mills ratio and its square. The two coefficients are oppositely signed and the impact on probability declines as the Inverse Mills ratio increases. There is some suggestion it

---

<sup>3</sup> This may be intuitively viewed as a continuous histogram.

turns up for high values of the inverse Mills ratio, but there are relatively few such values and this is probably the consequence of using a quadratic form. Column 4 shows the results of using the log of the inverse Mills ratio. The log-likelihood suggests this is superior to the simple linear form. In column 5 we replace the inverse Mills ratio and its square with  $\Phi(-\mathbf{Z}_i'\boldsymbol{\gamma})$ , the predicted probability of not using the Internet, and  $\Phi(-\mathbf{Z}_i'\boldsymbol{\gamma})^2$ . The results are similar, i.e. both significant at the 1% level and oppositely signed. The signs again suggest that the probability of using Internet banking declines, at a declining rate, as the marginality of being an Internet user increases. The same is true if we simply use  $\mathbf{Z}_i'\boldsymbol{\gamma}$  and its square<sup>4</sup>.

*Insert Figure 1 about here.*

The remaining columns show the results for the other two years. They are in essence similar to those already discussed. In the first of these we add business trust to the Internet banking equation which is significant and suggests a lack of trust in business reduces the probability of using Internet banking. This was the case in the original analysis by Orviska and Hudson. In the final set of results we replace this with trust in banks, which is again significant at the 1% level with the same implications. In both these cases the square of the Inverse Mills ratio is negatively significant at the 1% level, thus moderating the negative impact of an increase in the marginality of the Internet user. The resulting impacts are similar to that in Figure 1 and shown in Figures 2 and 3. However, there have been significant changes from the regressions with just the inverse Mills ratio included. In the second set of results the variable ‘town’ is now significant in the regressions with the square and log of the Inverse Mills ratio. In the final set of regressions ‘town’ moves from being significantly negative to insignificant and ‘village’ changes from being almost significantly negative to significantly positive in the final column. This is not surprising. The inverse Mills ratio is highly correlated with the right hand side variables in the selection equation, including the

---

<sup>4</sup> This contrasts with Vella (1998) where, in the context of a wage equation, the signs on  $\mathbf{Z}_i'\boldsymbol{\gamma}$  and  $(\mathbf{Z}_i'\boldsymbol{\gamma})^2$  were both positive and the latter insignificant.

locational variables. If we change the manner in which it is entered into the second stage regression, then it is to be expected that there may be some changes with respect to the other explanatory variables.

*Insert Figures 2 and 3 about here.*

The regressions relating to the selection equations, from which the inverse Mills ratios are derived, are shown in Table 2. There is very little difference between those estimated by the heckprob command and those estimated by probit. There is also a large degree of stability in the coefficients across the years, despite there being some changes in the included variables. The major difference appears to relate to age. In the earliest equation, Internet usage first increases, but then soon begins to decline sharply with age. Thus in all equations the probability of Internet usage declines with age for people older than 21 years.

*Insert Table 2 about here.*

One further point can be made with respect to these regressions. The full impact on Internet banking of a variable which is also present in the selection equation includes an indirect effect via its impact on the inverse Mills ratio. This is most easily illustrated with a regressions which include  $\Phi(-Z_i'\gamma)$  and  $\Phi(-Z_i'\gamma)^2$  rather than the inverse Mills ratio. The full impact of  $Z_k$ , given that the individual has access to the Internet, which appears in both equations, is then given by:

$$dY_i/dZ_{ik} = \beta_k - [\beta_\Phi \Phi(-Z_i'\gamma)\gamma_k + 2\beta_{\Phi^2} \Phi(-Z_i'\gamma)\Phi(-Z_i'\gamma)\gamma_k] \quad (1)$$

Where  $Y_i$  is the predicted value from the second stage regression,  $\beta_k$  is the coefficient on  $Z_{ik}$  in this second stage equation,  $\beta_\Phi$  and  $\beta_{\Phi^2}$  are the coefficients on  $\Phi(-Z_i'\gamma)$  and  $\Phi(-Z_i'\gamma)^2$

respectively and  $\gamma_k$ , equal to  $\partial Z_i' \gamma / \partial Z_{ik}$ , the coefficient on  $Z_{ik}$  in the selection equation. Hence For the variable ‘town’ in (1.10) and (2.4) this equals:

$$\begin{aligned} dY/dZ_{ik} &= 0.0675 + 1.820\phi(-Z_i' \gamma)(-0.1264) - 1.56\phi(-Z_i' \gamma)\Phi(-Z_i' \gamma)(-0.1264) \\ &= 0.0675 - 0.230\phi(-Z_i' \gamma) + 0.197\phi(-Z_i' \gamma)\Phi(-Z_i' \gamma) \end{aligned} \quad (2)$$

If we take the mid point of the distribution,  $\Phi(0)=0.5$  and  $\phi(0)=0.3989$  and  $dY/dZ_{ik} = -0.055$ . This is because town enters the selection equation with the effect that in general town dwellers are marginal Internet users, compared to city or large town dwellers. This impacts on the inverse Mills ratio in such a manner that the indirect effect is such that town dwellers are relatively less likely to use Internet banking. The combined effect of town, this indirect one coupled with that on town in the second stage equation ( $\beta_k$ ) is such as to reduce the initial positive impact of town dwelling on Internet banking usage. This impact depends upon  $\phi(-Z_i' \gamma)$  and  $\Phi(-Z_i' \gamma)$ . The greatest secondary impact of a variable is when  $-Z_i' \gamma$  is close to 0 and we are at the mid-point of the normal distribution. The secondary impact tails off slightly more for low marginal Internet users compared with the more marginal Internet user. In both cases the term in [.] in (1) approaches zero and the combined impact approaches  $\beta_k$ . This is shown in Figure 4.

It is important to emphasise that a variable which appears in both equations has three impacts on the probability of using Internet banking. Firstly, via the probability of being an Internet user and thus included in the sample for the second stage equation, secondly directly via this second stage equation and finally through the indirect effects in the second stage equation via the impact on Internet user marginality as reflected by the inverse Mills ratio or related variables. It is this latter effect we have emphasised and is relatively new to the literature.

*Insert Figure 4 about here.*



The assumption has been made that Internet user marginality, as reflected by the inverse Mills ratio, reflects confidence in the Internet. We now provide evidence for this assumption. In the Eurobarometer 74.3 survey there was a question which asked the extent to which the individual trusted Internet companies (search engines, social networking sites and email services) to protect their personal information. The Inverse mills ratio amongst Internet users who did trust was 0.442, but for those who did not trust it was 0.626. This difference is significant at the 1% level. Hence in this case, trust is clearly linked in the manner we have assumed with the inverse Mills ratio.

*Insert Table 2 about here.*

## 6 Conclusions

Our analysis suggests that marginal Internet users are less likely to use Internet banking, as in the standard Heckman equation, but that the impact is nonlinear. That is as individuals become steadily more marginal in their usage of the Internet, then the adverse impact on Internet banking usage flattens out<sup>5</sup>. This was our expectation when we began the work. It is often the case that the nonlinear impact of continuous variables is of this form with positive and second derivatives oppositely signed. We are also of the belief that this approach may be applicable for other applications of the Heckman methodology. In the Mincer curve, e.g., marginal labour force entrants may be marginal because they have a low reservation wage. In this case we would expect such marginality to be associated with lower earnings.

In terms of the policy implications, the results imply that the banks can increase the usage of Internet banking by increasing confidence in their own Web sites, which can be done by increasing the security of their Web site and then signalling this to potential users. But that banks can only do so much to instil confidence in their customers with respect to Internet banking. There also has to be confidence in the medium of the Internet itself. Governments

---

<sup>5</sup> This interpretation of our results is further confirmed by the results when we replace the Inverse Mills ratio with  $\Phi(Z_i'\gamma)$  or  $Z_i'\gamma$  as noted above.

can impact on this via regulation and the data suggests this is particularly important in the countries of Southern Europe. The speed and dependability of Internet access has also been highlighted in the literature as factors which may impact upon the adoption of online banking and may indeed impact on the marginal Internet user. This may well be resolved over time. The EU has set a target for all Europeans of 30 mbps (megabytes per second) by 2020 and if achieved should remove speed as a factor behind the take up of Internet banking. However despite this commitment, there must be doubts that remote communities will still lag behind others.

## References

- Aldas-Manzano J, Ruiz-Mafe C, Sanz-Blas S, Lassala-Navarré C (2011) Internet banking loyalty: evaluating the role of trust, satisfaction, perceived risk and frequency of use. *The Service Industries Journal* 31: 1165-1190.
- Bughin J (2003) The diffusion of Internet banking in Western Europe. *Electronic Markets* 13: 251-258.
- Cheng TCE, Lam DYC, Yeung ACL (2006) Adoption of Internet banking: an empirical study in Hong Kong. *Design Support Systems* 42: 1558-1572.
- Chinn M, Fairlie R (2006) The determinants of the global digital divide: A cross-country analysis of computer and Internet penetration. *Oxford Economic Papers* 59: 16-44.
- Courchane M, Nickerson D, Sullivan RJ (2002) Financial innovation, real options and endogenous competition – theory and applications to Internet banking. Paper presented at conference on Innovation in Financial Services and Payments, Federal Reserve Bank of Philadelphia.
- Cox J, Dale BG (2001) Service quality and e-commerce: An exploratory analysis. *Managing Service Quality* 11: 121-131.
- Cunningham LF, Gerlach J, Harper MD (2005) Perceived risk and e-banking services: An analysis from the perspective of the consumer. *Journal of Financial Services Marketing* 10: 165-78.
- DeYoung R, Lang WW, Nolle DE (2007) How the Internet affects output and performance at community banks. *Journal of Banking and Finance* 31: 1033-1060.
- Di Maggio P, Hargittai E, Neuman WR, Robinson JP, (2001) Social implications of the Internet. *Annual Review of Sociology* 27: 307-336.
- Dimitriadis S, Kyrezi, N (2011) The effect of trust, channel technology, and transaction type on the adoption of self-service bank channels. *The Service Industries Journal* 31:1293-1310
- European Commission (2010) A digital agenda for Europe. Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions. COM(2010) 245 final. 19.05.2010. Brussels.
- Frances PH, Paap R (2004) *Quantitative Models in Marketing Research*. Cambridge University Press, Cambridge
- Furst K, Lang WW, Nolle DE (2002) Internet banking. *Journal of Financial Services Research* 22: 93-117
- Gerrard P, Cunningham JB, Devlin JF (2006) Why consumers are not using Internet banking: A qualitative study. *Journal of Services Marketing*, 20, 160-8.
- Goldfarb A, Prince J (2008) Internet adoption and usage patterns are different: Implications for the digital divide. *Information Economics and Policy* 20: 2-15.

- Guerrero MM, Egea JMO, Gonzales MVR (2007) Application of the latent class regression methodology to the analysis of Internet use for banking transactions in the European Union. *Journal of Business Research* 60: 137-145.
- Hargittai E (1999) Weaving the western Web: explaining differences in Internet connectivity among OECD countries. *Telecommunications Policy* 23: 701-718.
- Heckman J (1976) The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models," *Annals of Economic and Social Measurement* 5: 475-492.
- Heckman JJ (1979) Sample Selection Bias, as a Specification Error. *Econometrica* 47: 153-161.
- Howcroft B, Hamilton R, Hewer P (2002) Consumer attitude and the usage and adoption of home-based banking in the United Kingdom. *The International Journal of Bank Marketing* 20: 111-121.
- Jianakoplos NA, Bernasek A (1998) Are women more risk averse? *Economic Inquiry* 36: 620-630.
- Johnston J, DiNardo J (1997) *Econometric Methods*, (New York: McGraw Hill, 1997).
- Kai L, Nagpurnanand PR (2007) Self-selection models in corporate finance. In: Eckbo BE (ed.) *Handbook of corporate finance - empirical corporate finance*. Elsevier Science, Amsterdam
- Kelton K, Fleischmann KR, Wallace WA (2008) Trust in digital information. *Journal of the American Society for Information Science and Technology* 59: 363-374.
- Kiiski S, Pohjola M (2002) Cross-country diffusion of the Internet. *Information Economics and Policy* 14: 297-310.
- Kontos, EZ, Bennett GG, Viswanath K (2007) Barriers and facilitators to home computer and Internet use among urban novice computer users of low socioeconomic position. *Journal of Medical Internet research* 9: e31.
- LaRose R, Gregg JL, Strover S, Straubhaar J, Carpenter S (2007) Closing the rural broadband gap: promoting adoption of the Internet in rural America. *Telecommunications Policy* 31: 359-373.
- Lee E-J, Eastwood DB, Lee J (2004) A sample selection model of consumer adoption of computer banking. *Journal of Financial Services Research* 26: 263-275.
- Lee MKO, Turban E (2001) A trust model for consumer internet shopping. *International Journal of Electronic Commerce* 6: 75-91.
- McKnight DH, Choudhury V, Kacmar C (2002) Developing and validating trust measures for e-commerce: An integrative typology. *Information Systems Research* 13: 334-359.
- Mukherjee A, Nath P (2003) A model of trust in online relationship banking. *International Journal of Bank Marketing* 21: 5-15.
- Orviska M, Hudson J (2009) Dividing or uniting Europe? Internet usage in the EU. *Information and Economic Policy* 21: 279-290.
- Pagan A, Vella F (1989) Diagnostic tests for models based on individual data: A survey, *Journal of Applied Econometrics* 4: S29-S59.
- Poon WC (2008) Adoption of e-banking services: the Malaysian perspective. *Journal of Business and Industrial Marketing* 23: 59-69.
- Recabarren M, Nussbaum M, Leiva C, (2007) Cultural illiteracy and the Internet. *Cyberpsychology and Behavior* 10: 853-856.
- Sinai T, Waldfogel J, (2004) Geography and the Internet: is the Internet a substitute or a complement for cities. *Journal of Urban Economics* 56: 1-24.
- Sullivan R, Wang Z (2005) Internet banking an exploration in technology diffusion and impact. Working Paper 05-05, Payments Systems Research Department, Federal Reserve Bank of Kansas City.
- Vella F (1998) Estimating Models with sample selection bias: A survey. *The Journal of Human Resources*, 33: 127-169.
- Wang Y, Lin H, Tang T (2003) Determinants of user acceptance of Internet banking: An empirical study. *International Journal of Service Industry Management*, 14(5), 501-519.

White H (1980) A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica* 48: 817–838.

## APPENDIX: DATA DEFINITIONS

Internet Access	A binary variable coded 1 if the individual responded that they used the Internet
Internet Banking	Coded 1 if the respondent indicated they used the Internet to access Internet banking ‘at least once a week’ (2005) or if they used the Internet for Internet banking (2004) or home banking (2010) with no time frame.
Male	male=1, female=0.
Age	The age, in years, of the respondent.
Self employed	Coded 1 if the individual was self-employed, otherwise 0.
Village	Coded 1 if the respondent lives in a rural area or village, otherwise 0.
Town	Coded 1 if the respondent lives in a small or middle sized town, otherwise 0.
Married	Coded 1 if the respondent is either married or living with partner, otherwise 0.
Manual worker	Coded 1 if the respondent is a manual worker, otherwise 0.
Unemployed	Coded 1 if the respondent is unemployed, otherwise 0.
Education	Age at which the individual finished full time education Coded: 1, <16 years; 2 16-19 years; 3 >19 years.
Pay bills	Difficulties in paying bills in the previous year, responses ranged from ‘most of the time’ (coded 1) to ‘almost never’ (coded 3).
Phone access	Coded 1 if access to fixed telephone services was difficult or ‘no access’, otherwise 0, (2004 data set).
Trust business	Coded 1 if the individual responded they trusted big companies, otherwise 0 and this option includes those who distrust them and also those who ‘did not know’, (2005 data set)
Trust banks	Coded 1 if the individual responded they trusted banks and financial institutions to collect and store personal information otherwise 0 and this option includes those who distrust them and also those who ‘did not know’, (2010 data set)
Trust Government	As for ‘trust banks’ but with respect to national public authorities. (2010 data set)

Notes: 3 data sets are used: Eurobarometer 74.3/Special Eurobarometer 359 November-December 2010, Eurobarometer 64.2 October-November 2005 and Eurobarometer 62.1 October-December 2004. Unless otherwise stated variables relate to all three surveys.

Table 1: Regression Results with Internet banking as the dependent variable.

	Eurobarometer 62.1					Eurobarometer 64.2					Eurobarometer 74.3				
	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	1.10	1.11	1.12	1.13	1.14	1.15
Male	0.1224** (4.33)	0.1287** (4.36)	0.1308** (4.42)	0.1297** (4.39)	0.1304** (4.41)	0.0921** (4.06)	0.102** (4.08)	0.1031** (4.13)	0.1006** (4.03)	0.1034** (4.14)	0.094** (4.36)	0.101** (4.49)	0.1088** (4.81)	0.1115** (4.91)	0.111** (4.91)
Age	0.0398** (6.41)	0.0409** (6.52)	0.0441** (7.11)	0.0434** (7.09)	0.0434** (7.03)	0.0601** (13.14)	0.0664** (13.45)	0.0707** (14.14)	0.0751** (15.52)	0.0725** (14.64)	0.0432** (10.86)	0.0442** (10.32)	0.0558** (13.14)	0.0595** (14.24)	0.05654** (13.31)
Age <sup>2</sup>	-0.0407** (5.45)	-0.0413** (5.40)	-0.0439** (5.84)	-0.0427** (5.79)	-0.0429** (5.76)	-0.0584** (10.97)	-0.0641** (11.15)	-0.0677** (11.79)	-0.0732** (13.51)	-0.0702** (12.46)	-0.0448** (9.97)	-0.0428** (8.77)	-0.0514** (10.82)	-0.0509** (11.17)	-0.05259** (11.14)
Village	0.0086 (0.23)	0.0146 (0.37)	0.0447 (1.12)	0.0534 (1.33)	0.0448 (1.12)	0.0093 (0.30)	0.0185 (0.55)	0.0374 (1.10)	0.0545 (1.57)	0.034 (1.00)	-0.0523 (1.83)	-0.0342 (1.16)	0.0057 (0.19)	0.0693* (2.28)	0.00497 (0.17)
Town	0.036 (1.02)	0.0404 (1.11)	0.057 (1.55)	0.0636 (1.72)	0.0567 (1.54)	0.0502 (1.74)	0.0579 (1.85)	0.0685* (2.18)	0.0801* (2.52)	0.0675* (2.15)	-0.0628* (2.30)	-0.0565* (2.01)	-0.0369 (1.31)	0.0052 (0.18)	-0.03682 (1.30)
Self-employed	0.2767** (5.75)	0.288** (5.71)	0.286** (5.66)	0.2836** (5.61)	0.286** (5.66)	0.1223** (3.19)	0.1336** (3.11)	0.1306** (3.03)	0.1415** (3.30)	0.1322** (3.07)	0.2512** (6.42)	0.2532** (6.27)	0.2391** (5.86)	0.2479** (6.13)	0.2406** (5.89)
Trust business						0.0995** (4.23)	0.1097** (4.25)	0.1103** (4.27)	0.115** (4.45)	0.1116** (4.32)					
Trust banks											0.2486** (9.79)	0.2577** (9.61)	0.2523** (9.37)	0.246** (9.16)	0.2538** (9.44)
Inverse Mills ratio		-0.4389** (7.47)	-1.036** (8.03)					-0.7543** (13.85)	-1.151** (11.31)			-0.6849** (13.09)	-1.585** (16.68)		
Inverse Mills ratio <sup>2</sup>			0.3157** (5.20)					0.2311** (5.07)					0.5804** (11.32)		
Log Inverse Mills ratio				-0.3132** (9.21)					-0.3555** (14.05)					-0.2777** (20.15)	
$\Phi(-\mathbf{Z}_i'\boldsymbol{\gamma})$					-1.596** (5.84)					-1.820** (8.22)					-2.816** (14.64)
$\Phi(-\mathbf{Z}_i'\boldsymbol{\gamma})^2$					0.7814** (2.83)					0.602** (2.58)					2.082** (9.25)
Observations	234468740			8740	8740	8740	27624	12687	12687	12687	12687	24779	15457	15457	15457
Log Likelihood	-15615			-5050	-5038	-5039	-19912	-6915	-6908	-6917	-6910	-17853	-8435	-8377	-8316
X <sup>2</sup>	852.1			1483	1534	1534	1192	1990	2076	2065	2081	2459	3317	3465	3366

Notes: Equations 1.1, 1.6 and 1.11 were estimated jointly with the selection equation shown in Table 2 using the heckprob command in STATA. The remaining equations were estimated using probit where the inverse Mills ratios were calculated from a first stage probit estimation also shown in Table 2.  $\Phi(-\mathbf{Z}_i'\boldsymbol{\gamma})$  represents the probability of not being an Internet user as estimated in the selection equation. Standard errors have been corrected for heteroscedasticity. (.) denotes t statistics, \*\*/\* significance at the 1%/5%

Table 2: Selection Equations; dependent variable is Internet access.

Eurobarometer:	62.1		64.2		74.3	
	2.1	2.2	2.3	2.4	2.5	2.6
Married	0.116** (5.33)	0.1037** (4.71)	0.129** (6.68)	0.1119** (5.62)	0.2362** (10.53)	0.2177** (9.54)
Education	0.596** (37.6)	0.5999** (37.77)	0.6038** (45.33)	0.6049** (44.88)	0.7021** (41.48)	0.6916** (40.12)
Age	0.0298** (7.57)	0.0305** (7.73)	-0.0147** (4.54)	-0.0147** (4.52)	-0.0142** (3.59)	-0.0133** (3.35)
Age <sup>2</sup>	-0.0714** (16.61)	-0.0721** (16.75)	-0.0217** (6.46)	-0.0219** (6.49)	-0.0399** (10.14)	-0.0417** (10.56)
Village	-0.2738** (10.47)	-0.273** (10.42)	-0.3088** (13.20)	-0.3084** (13.11)	-0.2945** (10.85)	-0.3041** (11.14)
Town	-0.1291** (5.04)	-0.1285** (5.01)	-0.1223** (5.30)	-0.1264** (5.45)	-0.1253** (4.60)	-0.1337** (4.88)
Unemployed	-0.7046** (17.53)	-0.6979** (17.20)	-0.4796** (13.73)	-0.4588** (12.78)	-0.5343** (14.51)	-0.5166** (13.82)
Manual worker	-0.4168** (16.78)	-0.4064** (16.18)	-0.4014** (18.16)	-0.386** (16.93)	-0.3716** (12.38)	-0.3526** (11.56)
Phone access	-0.362** (9.98)	-0.3545** (9.68)				
Trust Government					0.1293** (5.58)	0.1345** (5.71)
Pay bills					0.2311** (14.28)	0.2267** (13.74)
Observations	23446	23446	27624	27624	24779	24779
Log Likelihood	-15615	-10563	-19912	-12996	-17853	-9401
X <sup>2</sup>	852.1	9842	1192	12119	2459	14015

Notes: Equations 2.1, 2.3 and 2.5 were estimated jointly with the second stage equation shown in Table 1 using the heckprob command in STATA. The remaining equations were estimated using probit and then used to calculate the inverse Mills ratios used in Table 1. Standard errors have been corrected for heteroscedasticity. (.) denotes t statistics, \*\*/\* significance at the 1%/5% levels.



